



Assessment of spatial structure of groundwater quality variables based on the entropy theory

Y. Mogheir, J. L. M. P. de Lima, V. P. Singh

► To cite this version:

Y. Mogheir, J. L. M. P. de Lima, V. P. Singh. Assessment of spatial structure of groundwater quality variables based on the entropy theory. Hydrology and Earth System Sciences Discussions, 2003, 7 (5), pp.707-721. hal-00304920

HAL Id: hal-00304920

<https://hal.science/hal-00304920>

Submitted on 1 Jan 2003

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Assessment of spatial structure of groundwater quality variables based on the entropy theory

Y. Mogheir¹, J.L.M.P. de Lima² and V.P. Singh³

¹Department of Civil Engineering, Faculty of Science and Technology - Campus 2, University of Coimbra, 3030-290 Coimbra, Portugal

²Institute of Marine Research - Coimbra Interdisciplinary Centre; Department of Civil Engineering, University of Coimbra, 3030-290 Coimbra, Portugal

³Department of Civil and Environmental Engineering, Louisiana State University, Baton Rouge, LA 70803-6405, USA

Email for corresponding author: plima@dec.uc.pt

Abstract

Fundamental to the spatial sampling design of a groundwater quality monitoring network is the spatial structure of groundwater quality variables. The entropy theory presents an alternative approach to describe this variability. A case study is presented which used groundwater quality observations (13 years; 1987–2000) from groundwater domestic wells in the Gaza Strip, Palestine. The analyses of the spatial structure used the following variables: Electrical Conductivity (EC), Total Dissolved Solids (TDS), Calcium (Ca), Magnesium (Mg), Sodium (Na), Potassium (K), Chloride (Cl), Nitrate (NO₃), Sulphate (SO₄), alkalinity and hardness. For all these variables the spatial structure is described by means of Transinformation as a function of distance between wells (Transinformation Model) and correlation also as a function of distance (Correlation Model). The parameters of the Transinformation Model analysed were: (1) the initial value of the Transinformation; (2) the rate of information decay; (3) the minimum constant value; and (4) the distance at which the Transinformation Model reaches its minimum value. Exponential decay curves were fitted to both models. The Transinformation Model was found to be superior to the Correlation Model in representing the spatial variability (structure). The parameters of the Transinformation Model were different for some variables and similar for others. That leads to a reduction of the variables to be monitored and consequently reduces the cost of monitoring.

Keywords: transinformation, correlation, spatial structure, municipal wells, groundwater monitoring, entropy

Introduction

Prior to the design of groundwater quality monitoring networks, it is essential to investigate the spatial structure of the groundwater quality variables to be monitored. Generally, the objective of monitoring introduces these variables. For example, greater monitoring effort is required for groundwater to be used for domestic (municipal) purposes than for agricultural use. Several variables, such as EC, TDS, Cl, NO₃, bacteria, alkalinity, hardness, pH, etc., are usually measured in the groundwater used for municipal purposes. The aim of characterising the spatial structure of these variables is not only to assess their monitoring, which is needed to assess the cost of monitoring, but also to give a clear picture about their spatial variability (structure). The spatial structure of the groundwater quality variables can produce, for example, contour maps of the variable means. These maps can be used for predicting and signifying pollution areas. Accordingly, protection measures and

management and planning decisions can be made to minimise the deterioration in the polluted areas.

Groundwater quality variables can often be characterised by a known distribution in space, called the spatial distribution function; examples are: EC, TDS, Cl, NO₃, groundwater level, rainfall, etc. The spatial distribution function is known in the literature as the semivariogram (e.g. Journel and Huijbregts, 1978; Olea, 1999) which contains the structural information on the variable. In applications of the Kriging technique, the spatial correlation structure (experimental semivariogram or covariance) of the variable is determined from real data (measurements). Consequently, to complete the estimation approach using the Kriging technique, analytical curves, such as linear, spherical, exponential, Gaussian and cubic curves, are fitted to the experimental semivariogram (e.g. Marsily, 1986; Cressie, 1990). Kriging techniques are employed to assess the spatial dependencies of the water quality variables such

as TDS (e.g. Jager *et al.*, 1990), groundwater level (e.g. Hudak and Loaiciga, 1993; Geo *et al.*, 1996; Ahmed, 2002) and groundwater quality variables such as NO_3 and Cl (e.g., Rouhani and Hall, 1988; Loaiciga, 1989).

Application of the entropy theory in spatial structure analysis can be comparable to semivariogram analysis. Entropy theory has four types of entropy (also referred to as coefficients): marginal entropy, conditional entropy, joint entropy and Transinformation (e.g., Kapur and Kesavan, 1992; Lubbe, 1996). Transinformation was used to measure the dependency between different water quality variables (Wu and Zidek, 1992; Harmancioglu *et al.*, 1999), rainfall values measured at different points (e.g. Sounga, 1976; Husain, 1989; Krastanovic and Singh, 1992), groundwater level (e.g. Bueso *et al.*, 1998) and Cl (e.g. Mogheir and Singh, 2002a).

Mogheir and Singh (2002a) characterised the spatial structure of a groundwater quality variable by means of Transinformation as a function of distance (T Model) and joint entropy as a function of distance. In this article, the main objective is to investigate the ability of the T Model to characterise the spatial structure of different groundwater quality variables that are being monitored by the municipal monitoring wells in the Gaza Strip, Palestine. These variables are: Electrical Conductivity (EC), Total Dissolved Solids (TDS), Calcium (Ca), Magnesium (Mg), Sodium (Na), Potassium (K), Chloride (Cl), Nitrate (NO_3), Sulphate (SO_4), alkalinity and hardness. Electrical Conductivity (EC) is defined as the ability of a substance to conduct an electric current. In this respect, salt does not conduct when it is solid but becomes a conductor when it is dissolved in water. Therefore, EC is used to estimate the amount of total dissolved salts (TDS), or the total amount of dissolved ions in the water. Alkalinity is a measure of the concentration of carbonate, bicarbonate, and hydroxide, and contributes to the total hardness. Alkalinity is also defined as the ability to neutralise acids. The alkalinity is the total concentration of calcium carbonate (CaCO_3). Hardness is a measure of the concentration of calcium and magnesium salts in water. Hardness reacts chemically with soap and the higher the hardness the more soap is required to form lather. The alkalinity and hardness of groundwater supply generally depend on the soil or rocks from which the water is derived. High alkalinity and hardness in groundwater are derived from chalk and limestone in the aquifer. The units of EC are microSiemens per centimetre, $\mu\text{S cm}^{-1}$, while the units of the rest of groundwater quality variables are milligram per litre, mg l^{-1} (William, 1997).

The Transinformation Model (T Model) (e.g. Harmancioglu *et al.*, 1999; Mogheir and Singh, 2002a) of each of the previously mentioned variables will be

determined and will be compared with correlation as a function of distance between wells (Correlation Model, C Model) (e.g. Mogheir *et al.* 2004a, b). The T and C Models for all groundwater quality variables will be represented by the exponential decay curves. The T Model contains four parameters: (1) the initial value of Transinformation T_0 ; (2) the Transinformation decay rate (k); (3) the minimum Transinformation value T_{\min} ; (4) the distance at which the Transinformation Model approaches a constant minimum value (L). Using these parameters, the spatial variability of the groundwater quality variables will be assessed.

Groundwater in the Gaza Strip

GROUNDWATER QUANTITY

The Gaza Strip, located in the southern corner of Palestine, is divided into five Governorates: Northern, Gaza, Middle, Khanyunes and Rafah (Fig. 1). Each Governorate has its own municipal wells. The water resources in the Gaza Strip are essentially limited to the part of the groundwater in the coastal aquifer that underlies its 360 km^2 area. This coastal aquifer is the only aquifer in the Gaza Strip and is composed of Pleistocene marine sand and sandstone, intercalated with clay layers. The maximum thickness of different bearing horizons occurs in the northwest along the coast (150 m) and decreases gradually towards the east and south-east along the eastern border of the Gaza Strip to less than 10 m (Fig. 2). Near the coast, coastal clays extend about 2–5 km inland, and divide the aquifer sequence into three or four subaquifers, depending upon the location (referred to as subaquifers A, B1, B2, and C — Fig. 2). The base of the coastal aquifer is formed of impervious clay shale rocks of Neogene age (Saqiyah formation) with total thickness ranging between 500 to 1000 m. The depth to groundwater level of the coastal aquifer varies between a few metres in the low land area along the shoreline and about 70 m along the eastern border. The coastal aquifer holds approximately $5 \times 10^9 \text{ m}^3$ of groundwater of varying quality. However, only $1.4 \times 10^9 \text{ m}^3$ of this is 'freshwater', with Cl content of less than 500 mg l^{-1} . This fresh groundwater typically occurs in the form of lenses that float on the top of the brackish and/or saline groundwater. That means approximately 70% of the aquifer is brackish or saline water and only 30% is fresh water found mainly in the Northern Governorate. Moreover, the freshwater can be found in the shallow aquifer (A subaquifer — Fig. 2) where most of the groundwater wells (municipal and agricultural) penetrate. The major source of water that generates the renewable groundwater in the aquifer is rainfall. Rainfall is sporadic across the Gaza Strip and generally varies from 400 mm year^{-1} in the north to

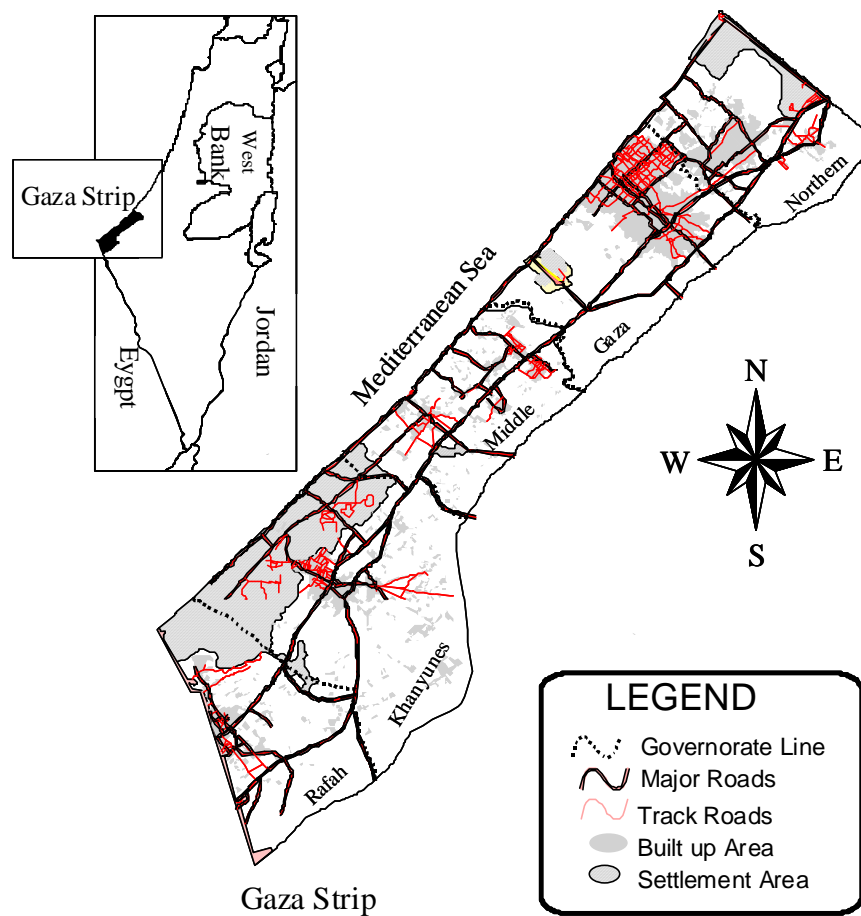


Fig. 1. Location map of Gaza Strip, Palestine.

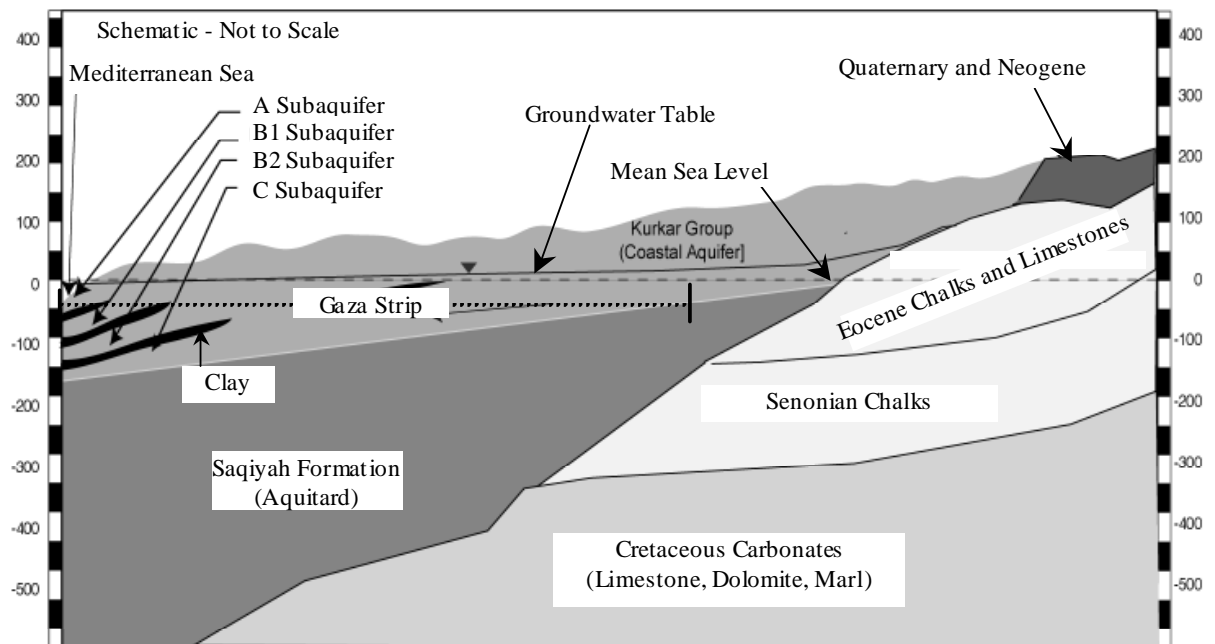


Fig. 2. Generalised hydrogeological cross section of the Gaza Strip aquifer (Source: Metcalf and Eddy, 2000).

about 200 mm year⁻¹ in the south (Metcalf and Eddy, 2000).

The lateral inflow to the aquifer is estimated at between $10 \times 10^6 \text{ m}^3 \text{ year}^{-1}$ and $15 \times 10^6 \text{ m}^3 \text{ year}^{-1}$. Some recharge is available from the major surface flow (Wadi Gaza). However, the extensive extraction from Wadi Gaza in Israel limits this recharge to 1.5×10^6 to $2 \times 10^6 \text{ m}^3$. As a result, the total freshwater recharge at present is limited to approximately 56×10^6 to $62 \times 10^6 \text{ m}^3 \text{ year}^{-1}$ (Metcalf and Eddy, 2000). Under natural conditions, groundwater flow in the Gaza Strip is towards the Mediterranean Sea, where it discharges. However, pumping over 40 years has disturbed natural flow patterns significantly. In the year 2000, a large cone of depression has formed in the northern and southern part of the Gaza Strip, where the groundwater levels are approximately 2 m below the mean sea level.

GROUNDWATER QUALITY

More water has been pumped from the aquifer than recovered by the aquifer in recent decades. This over-extraction from the aquifer has resulted in a significant drawdown of the groundwater, with resulting intrusion of seawater and upconing of the underlying saline water. The major water quality problems are the high salinity (Cl) and the high nitrate (NO₃) concentrations in the aquifer. The Cl

and NO₃ concentrations in some of the municipal wells in 1999 are shown in Fig. 3. High levels of Cl groundwater have caused high salinity in the water supply. Most of the municipal wells exceed the WHO drinking standard of Cl (250 mg l⁻¹). In addition, most of the municipal wells in the Gaza Strip show nitrate levels in excess of the WHO drinking water standard of 50 mg l⁻¹. In urban areas the main sources of nitrates are domestic sewage effluent and in rural areas the problem in the use of fertilisers and its concentration is increasing at a rate up to 10 mg l⁻¹ per year. As shown in Fig. 3, if both Cl and NO₃ are considered, the potable groundwater is found in the northern part of the Gaza Strip. Therefore, considerable monitoring and control measures are envisaged by the Palestinian Water Authority (PWA) in this part of the Gaza Strip.

GROUNDWATER MONITORING NETWORKS

In the Gaza Strip, the groundwater monitoring networks are divided into: (1) groundwater quality, (2) groundwater depth (level) and (3) municipal wells. The groundwater quality network measures three variables: EC, Cl and NO₃. The measurements are undertaken twice per year using 400 agricultural wells (Mogheir and Singh, 2002b). The groundwater depth network measures the groundwater depth

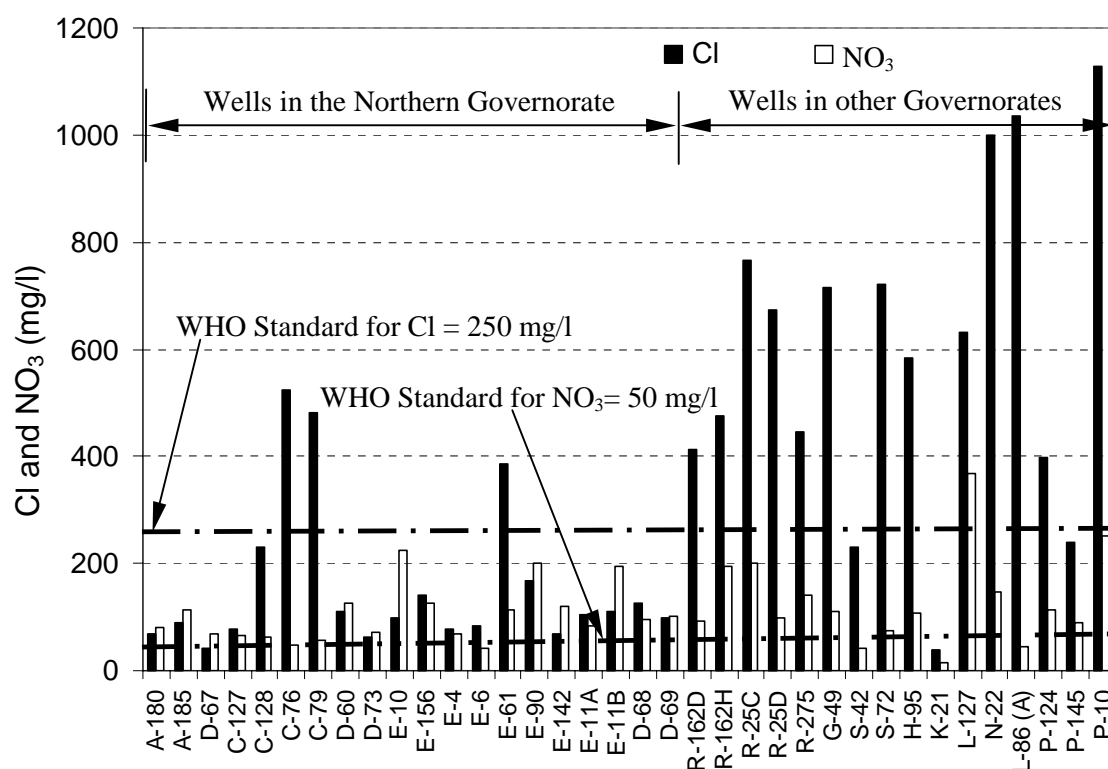


Fig. 3. Cl and NO₃ concentration in some of the domestic municipal wells in the Gaza Strip, in 2000. The location of A-180, A-185, D-67, C-127, C-128, E-4, E-90, E-11A and E-11B is given in Fig. 4.

on a monthly basis, using around 120 agricultural wells. The third type of monitoring network consists of the municipal wells which measure: EC, TDS, Ca, Mg, Na, K, Cl, NO₃, SO₄, alkalinity and hardness. For this purpose, around 100 municipal wells are used. The measurements are carried out twice per year. The three networks are used to collect information that describes the groundwater quality and quantity status.

Most of the municipal wells were drilled after 1997 and, therefore, they only have a few groundwater quality measurements. In this paper, the groundwater quality measurements from 13 municipal wells in the Northern Governorate will be analysed. These wells are drilled in A Subaquifer of the total Gaza Strip aquifer (Fig. 2), which can be assumed homogeneous. The wells measure 11 groundwater quality variables: EC, TDS, Ca, Mg, Na, K, Cl, NO₃, SO₄, alkalinity and hardness. The variables were measured twice per year: winter and summer, in the period

1987 to 2000. The location of these 13 wells in the Northern Governorate is shown in Fig. 4. The missing data were interpolated linearly for all the variables. The spatial variations of the mean of the variables are presented in Fig. 5 where the contour lines were drawn using the Kriging technique, which is an option in Surfer-7 mapping program (Golden Software, 1999). The time series of some variables (Ca, Mg, Na, Cl, NO₃, SO₄) for wells D-67, E-11B and C-128 are illustrated in Fig. 6; the basic statistics of the time series of all variables are presented in Table 1.

Applied Methodology

MEASURING SPATIAL DEPENDENCY

The entropy of a random variable is a measure of information or uncertainty associated with it. For a random variable x , the marginal entropy, $H(x)$, can be defined as the potential

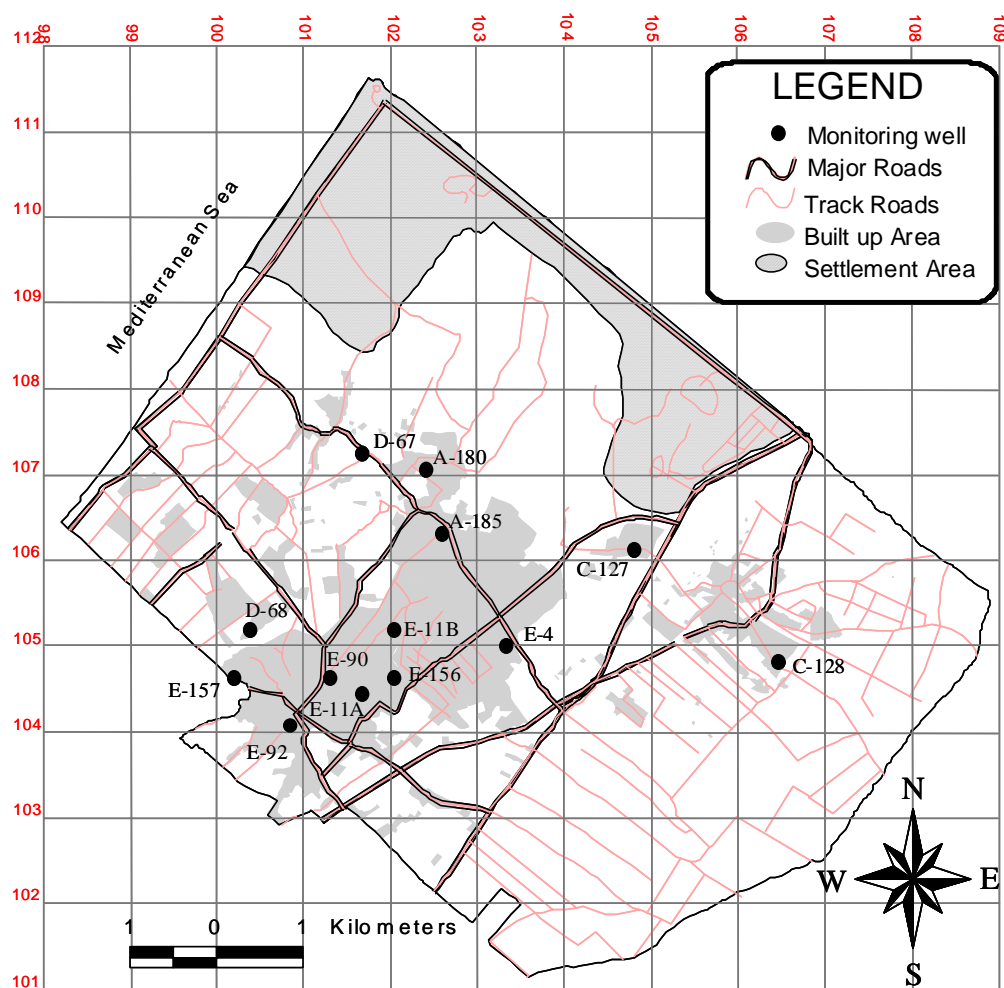


Fig. 4. Location of selected groundwater quality monitoring wells (●) in the Northern Governorate.

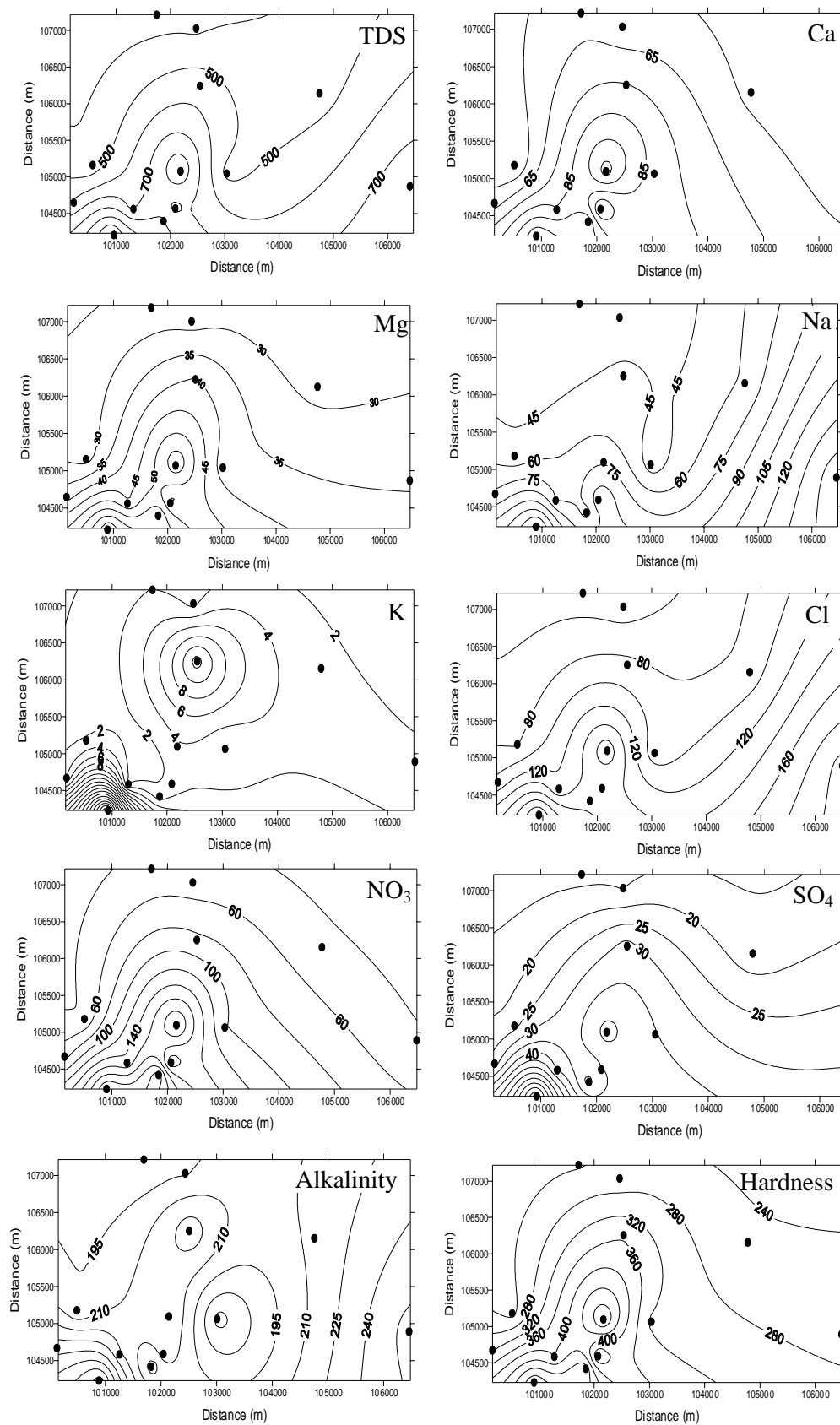


Fig. 5. Average concentration distribution map of: TDS, Ca, Mg, Na, K, Cl, NO₃, SO₄, alkalinity, and hardness in the Northern Governorate, for the year 2000.

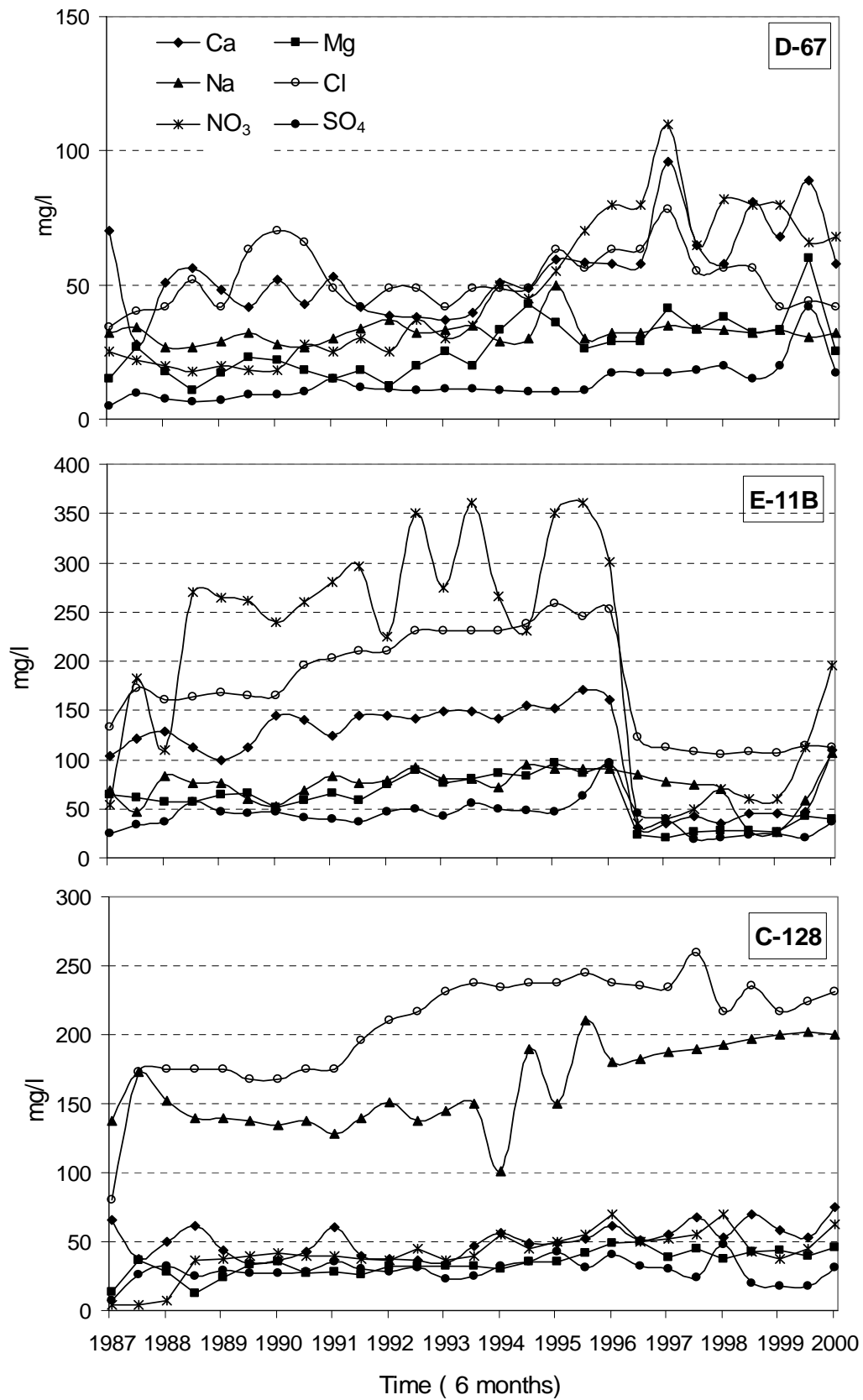


Fig. 6. Time series of variables: Ca, Mg, Na, Cl, NO₃ and SO₄ for municipal wells: D-67, E-11B and C-128. The time series were over the period 1987 to 2000.

Table 1. Mean (\bar{X}) and standard deviation (σ) of the groundwater variables (EC, TDS, Ca, Mg, Na, K, Cl, NO_3 , SO_4 , Alkalinity, Hardness) for the 13 monitoring wells (A-180, A-185, C-127, C-128, D-67, D-68, E-157, E-11A, E-11B, E-156, E-4, E-90, E-92) as presented in Fig. 4.

	A-180	A-185	C-127	C-128	D-67	D-68	E-157	E-11A	E-11B	E-156	E-4	E-90	E-92
EC ($\mu\text{S cm}^{-1}$)													
\bar{X}	561	841	722	1204	534	654	936	1262	1368	1037	723	1043	1746
σ	96	119	69	182	136	226	200	161	352	72	85	176	215
TDS (mg l^{-1})													
\bar{X}	383	556	460	782	358	425	597	813	895	676	475	679	1115
σ	52	70	56	99	75	120	129	81	253	51	58	98	120
Ca (mg l^{-1})													
\bar{X}	59	75	52	51	62	45	61	103	110	62	81	86	114
σ	15	18	9	12	47	20	15	18	46	11	105	14	13
Mg (mg l^{-1})													
\bar{X}	28	40	28	34	27	28	38	59	59	44	37	45	70
σ	6	8	5	9	11	7	6	12	24	6	7	9	15
Na (mg l^{-1})													
\bar{X}	31	50	63	162	32	51	78	69	74	104	38	74	157
σ	7	9	10	29	4	26	29	23	19	11	15	19	37
K (mg l^{-1})													
\bar{X}	1,4	13,0	2,7	2,3	2,0	2,3	3,7	3,1	3,8	2,9	2,1	2,3	38,9
σ	0,5	3,0	1,4	0,5	0,5	0,5	1,0	0,7	1,1	1,0	0,8	0,9	20,5
Cl (mg l^{-1})													
\bar{X}	51	86	88	207	52	78	106	154	176	139	90	123	224
σ	8	13	12	38	11	30	37	15	53	12	13	22	32
NO_3 (mg l^{-1})													
\bar{X}	51	85	46	42	47	44	74	232	206	125	101	146	222
σ	15	28	19	16	27	23	17	76	111	23	41	34	58
SO_4 (mg l^{-1})													
\bar{X}	15	32	17	29	13	23	27	28	42	34	35	40	80
σ	7	10	8	8	7	12	19	4	16	49	11	10	15
Alkalinity (mg l^{-1})													
\bar{X}	194	236	214	261	180	197	249	186	225	220	159	226	307
σ	27	21	22	37	39	25	50	34	26	13	49	25	22
Hardness (mg l^{-1})													
\bar{X}	258	355	249	268	236	228	319	513	523	319	311	401	559
σ	42	55	21	54	57	70	53	71	207	65	37	69	55

information of the variable. The dependency (mutual information) between two variables can be measured by conditional entropy, joint entropy and Transinformation. For two random variables, x and y , the conditional entropy $H(x|y)$ is a measure of the information content of x which is not contained in the random variable y . The joint entropy, $H(x, y)$, is the total information content contained in both x and y . The mutual entropy (information) between x and y ,

also called Transinformation, $T(x, y)$, is interpreted as the reduction in uncertainty in x , due to the knowledge of the random variable y . It can also be defined as the information content of x that is contained in y . The Transinformation for discrete variables can be expressed as (e.g. Mogheir and Singh, 2002a)

$$T(x, y) = - \sum_{i=1}^n \sum_{j=1}^m p(x_i, y_j) \ln \left[\frac{p(x_i, y_j)}{p(x_i)p(y_j)} \right] \quad (1)$$

where x and y are two discrete variables with values x_i , $i=1, 2, \dots, n$; y_j , $j=1, 2, \dots, m$, defined in the same probability space, each of which has a discrete probability of occurrence $p(x_i)$, $p(x_i, y_j)$ is the joint probability of x_i , y_j and $p(x_i|y_j)$ is the probability of x_i conditional on y_j . Note that $H(x, y) = H(y, x)$.

Many entropy theory-based applications have computed Transinformation as a function of correlation between variables. This approach presumes knowledge of the probability distributions of the random variables under study (e.g. Harmancioglu *et al.*, 1999). The problem of not knowing the probability distributions can, however, be circumvented if a discrete approach is adopted (e.g. Mogheir *et al.*, 2003a). To calculate the discrete Transinformation values, the joint or conditional probability is needed, and this can be obtained using a contingency table (e.g. Gokhale and Kullback, 1978; Mogheir and Singh, 2002a).

Contingency tables (cross-tabulations) record the frequency for the values that fall into each possible combination of two categories. To construct a two dimensional contingency table (Table 2), the following steps are involved (e.g. Mogheir *et al.*, 2003b).

- Let the random variable x have values consisting of v categories (class intervals), while the random variable y is assumed to have u categories (class intervals). There is no general rule for choosing the number of class intervals (NCI). To avoid such complications the number of class intervals was calculated as (e.g., Zhou, 1996):

$$NCI = 1 + 1.33 \ln(n) \quad (2)$$

where, NCI is the number of class intervals and n is the size of the time series of the variable. Notice that, the number of class intervals should be the same for all the

time series of the groundwater quality variables.

- The cell density or the joint frequency for (i, j) is denoted by f_{ij} , $i = 1, 2, \dots, v$; $j = 1, 2, \dots, u$, where the first subscript refers to the row and the second subscript to the column. The cell density is the count of measurements with the corresponding u class interval of variable y and v class interval of variable x .
- The marginal frequencies are denoted by f_i and f_j for the row and the column values of the variables, respectively. f_i is the summation of the cell density for each v categories of variable x .
- The joint probability $p(x_i, y_j)$ is computed by dividing the cell density by the total of the data recorded in one well.

A dependency can also be measured by means of Correlation or Covariance between variables (Olea, 1999). The Correlation coefficient r_{xy} was calculated for each pair of wells as:

$$r_{xy} = \frac{\text{Cov}_{xy}}{S_x S_y} \quad (3)$$

where Cov_{xy} is the Covariance between random variables x and y ; S_x and S_y are the standard deviations of variables x and y , respectively Cov_{xy} can be obtained as

$$\text{Cov}_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n - 1} \quad (4)$$

where \bar{x} and \bar{y} are the means of variable x and y , respectively.

The spatial dependency was measured by the Transinformation and Correlation Models. These models were determined by computing Transinformation and Correlation, and the distance between wells.

EXPONENTIAL DECAY FITTING APPROACH

Nonlinear decays are common in physical systems. One of these decays used to model many physical processes is the exponential decay. It is used whenever the rate at which something happens is proportional to the amount that is left. An example of the application of the exponential decay is the radioactive isotope decay. In this process, the number of atoms that decay in any short interval is proportional to the number of undecayed atoms that were present at the beginning of the interval. The total decay of the sample decreases with time because there are fewer and fewer undecayed atoms (e.g. Motulsky, 1999a). Note, the atoms decay extremely rapidly at first, however, as the time increases, the decayed atoms fall ever more slowly. For this

Table 2. Two-dimensional contingency table (frequency).

		y					
		1	2	3	...	u	Total:
x	1	f_{11}	f_{12}	f_{13}	...	f_{1u}	$f_{1.}$
	2	f_{21}	f_{22}	f_{23}	...	f_{2u}	$f_{2.}$
	3	f_{31}	f_{32}	f_{33}	...	f_{3u}	$f_{3.}$

	V	f_{v1}	f_{v2}	f_{v3}	...	f_{vu}	$f_{v.}$
	Total:	$f_{.1}$	$f_{.2}$	$f_{.3}$...	$f_{.u}$	$f_{..}$ or f_y

reason, the half-life parameters are introduced as the time needed for the atoms to decay to half their original amount. With an exponential decay process, a long time is spent at the tail of the exponential decay curve (total life). The total life is the time required for atoms to decay completely. In atomic decay applications, the half-life has greater applications than the total life, due to the difficulty of estimating the total life amount and the interest is in the amount of decayed atoms.

The Transinformation as a function of distance between wells can be analogous to the case of atoms decaying as a function of time. In the exponential Transinformation decay curve, the initial Transinformation value is decayed over a certain distance, is depleted and reaches the minimum value at a distance equal to the range (L). In Transinformation applications, the objective is to determine this value of distance, which is comparable to the total life in the atomic decay case.

Mogheir *et al.* (2004b) suggested that the exponential decay curve could be the best representation of the discrete T Model, and could be presented as:

$$T(d) = G e^{(-Kd)} + Q \quad (5)$$

where the exponential decay curve starts with $T_0 = G + Q$ at distance (d) = 0; and the curve decays to reach a Q value with a constant rate K (1 m^{-1}). The units of G and Q are expressed in the same way as the T unit 'Nats' (e.g. Husain, 1989). Equation (4) was also applied to the Correlation Model (C Model).

The fitting of the exponential decay curve to the discrete models was performed using the least squares fitting procedure with the GRAPHPAD PRISM3 statistical software (Motulsky, 1999b). The coefficient of determination was used to quantify the goodness of fit between the exponential decay curve and discrete model data. The coefficient of determination (R^2) was computed as:

$$R^2 = 1.0 - \frac{SS_{\text{reg}}}{SS_{\text{tot}}} \quad (6)$$

where SS_{reg} is the sum of the squares of the residuals between the discrete model and the best fit exponential decay curve, and SS_{tot} is the sum of the squares of the residuals between the discrete model and the horizontal line through the mean.

DETERMINING L (THE RANGE) IN THE T MODEL

In assessing and redesigning a groundwater monitoring network, the definition of distance between monitoring wells (L) is particularly crucial. An accurate estimation of the

range L plays an important role in the comparison between characterisations of the spatial structures of different variables. In the literature, the range of the semivariogram is specified according to the threshold value of the sill. For example, Journel and Huijbregts (1978) took the range at 95% of the sill. Others recommend specifying the range of the semivariogram by hand drawing or by specifying different range values and making different scenarios (e.g. Olea, 1984; Yfantis *et al.*, 1987). Mogheir *et al.* (2004a) determined the value of L by drawing two lines, the first parallel to the declining part of the T Model and the second parallel to steady part of the T Model. The intersection point of the two lines indicated the L value.

In this study, L value will be determined as follows. The Transinformation as a function of distance (exponential decay curve) in Eqn. (4) can be written as:

$$T(d) = (T_0 - T_{\min})e^{(-Kd)} + T_{\min} \quad (7)$$

where T_0 is the initial value of Transinformation (K) is the Transinformation decay rate, T_{\min} the minimum Transinformation value and d is the distance between wells. This L value is estimated as the distance at which the Transinformation Model approached a constant minimum value T_{\min} . For this exponential decay curve, $T(d)$ asymptotically approaches T_{\min} , therefore, the L value is estimated when $T(d)$ is very close to T_{\min} . It is assumed that at $d = L$, $T(L) = T_{\min} + \varepsilon$, where ε is a constant number and its unit is Nats. If $T(L)$ is substituted in Eqn. (7) by $T_{\min} + \varepsilon$, Eqn. (7) can be written as

$$\begin{aligned} T_{\min} + \varepsilon &= (T_0 - T_{\min})e^{(-KL)} + T_{\min} \\ \varepsilon &= (T_0 - T_{\min})e^{(-KL)} \end{aligned} \quad (8)$$

Taking the logarithm of both side of Eqn. (8)

$$\ln(\varepsilon) = \ln(T_0 - T_{\min}) - KL \quad (9)$$

Then the value of L can be computed as

$$L = \frac{\ln(T_0 - T_{\min}) - \ln(\varepsilon)}{K} \quad (10)$$

From Eqn. (10), L is inversely proportional to K. When K increases, the value at which the T Model reaches the minimum value decreases. The value of L is also affected by ε . The higher ε value (difference between $T(L)$ and T_{\min}) to be considered, the lower L value will be estimated. The ε value can be expressed as a percentage from T_0 as

$$\% \varepsilon = \frac{\varepsilon}{T_0} \times 100 \quad (11)$$

where T_0 is the initial Transinformation value.

Assessment of the Spatial Structure

CORRELATION MODELS (C MODELS)

The C Models for all variables were obtained by computing the correlation coefficient using Equation (2) and the distance between wells. The exponential decay curve is fitted

to the C Models' data using the least squares technique. An example of fitting the exponential curve to the C Model data for the groundwater quality variable EC is shown in part (A) of Fig. 7. The C Model for each variable is presented in Fig. 8 and Table 3.

Table 3. Fitting correlation ($r(d)$, C Model) and Transinformation ($T(d)$, T Model) with exponential decay curves for the groundwater quality variables: EC, TDS, Ca, Mg, Na, K, Cl, NO_3 , SO_4 , alkalinity and hardness.

Water Quality Variables	Fitting Equation	R^2
EC	$r(d) = 0,6220 e^{(-0,0048 d)} + 0,3710$	0,46
	$T(d) = 0,9900 e^{(-0,0034 d)} + 0,4200$	0,77
TDS	$r(d) = 0,7700 e^{(-0,0023 d)} + 0,2201$	0,51
	$T(d) = 0,9370 e^{(-0,0024 d)} + 0,4200$	0,84
Ca	$r(d) = 0,8020 e^{(-0,0048 d)} + 0,1801$	0,49
	$T(d) = 0,8900 e^{(-0,0032 d)} + 0,3200$	0,63
Mg	$r(d) = 0,7932 e^{(-0,0043 d)} + 0,2100$	0,47
	$T(d) = 1,0500 e^{(-0,0050 d)} + 0,3200$	0,55
Na	$r(d) = 0,5907 e^{(-0,0021 d)} + 0,3800$	0,33
	$T(d) = 0,8500 e^{(-0,0060 d)} + 0,5100$	0,66
K	$r(d) = 0,7400 e^{(-0,0034 d)} + 0,2500$	0,30
	$T(d) = 0,7700 e^{(-0,0034 d)} + 0,3167$	0,74
Cl	$r(d) = 0,8801 e^{(-0,0020 d)} + 0,1108$	0,44
	$T(d) = 0,9600 e^{(-0,0028 d)} + 0,4100$	0,74
NO_3	$r(d) = 0,8802 e^{(-0,0052 d)} + 0,1806$	0,35
	$T(d) = 0,9500 e^{(-0,0051 d)} + 0,4186$	0,86
SO_4	$r(d) = 0,4920 e^{(-0,0032 d)} + 0,5004$	0,41
	$T(d) = 0,7800 e^{(-0,0034 d)} + 0,4186$	0,58
Alkalinity	$r(d) = 0,8100 e^{(-0,0006 d)} + 0,1108$	0,44
	$T(d) = 1,0500 e^{(-0,0034 d)} + 0,4600$	0,60
Hardness	$r(d) = 0,8000 e^{(-0,0020 d)} + 0,1007$	0,41
	$T(d) = 0,8800 e^{(-0,0034 d)} + 0,4550$	0,73

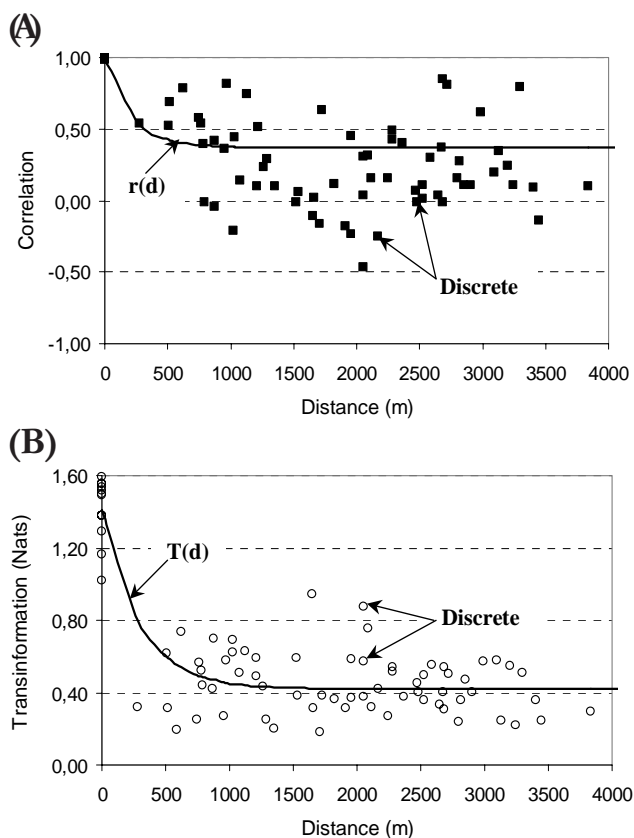


Fig. 7. (A) Correlation Models; and (B) Transinformation Models for the groundwater quality variable: EC.

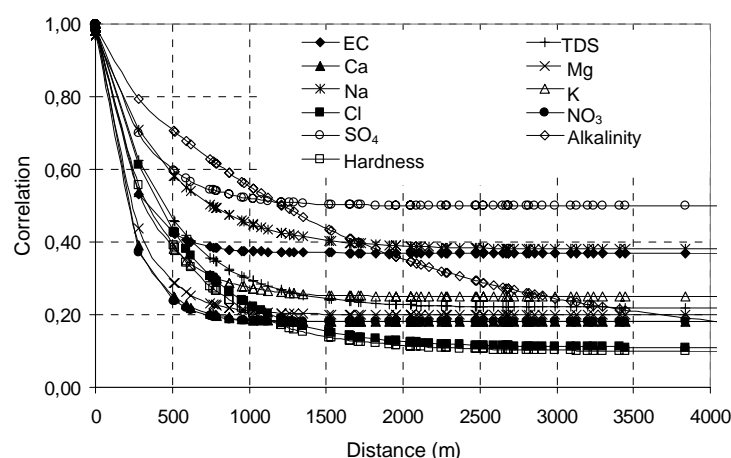


Fig. 8. Correlation Model for groundwater quality variables: EC, TDS, Ca, Mg, Na, K, Cl, NO_3 , SO_4 , alkalinity and hardness. The symbols (\bullet , \blacklozenge , \blacktriangle , etc.) are used to distinguish between the eleven C Models of the different groundwater quality variables.

TRANSFORMATION MODELS (T MODELS)

The T Models for the groundwater variables were obtained by computing the T values using Eqn. (1) and the distance between wells. The exponential decay curve is fitted to the T Model data using the least square technique (see part (B) of Fig. 7 for EC as an example). The T Model for each variable is presented in Fig. 9 and Table 3.

In Table 2, R^2 ranges between 0,55 to 0,85 in the T Models, where it ranges between 0,30 to 0,51 in the C Models. The variation of R^2 of C Models and R^2 of T Models are plotted in Fig. 10. In this figure, all the variable symbols lie in a small region close to the upper left corner of the graph. Hence, this figure and Table 2 show that, for all variables, the T Model selected as an exponential decay model infers spatial variability better than does the C Model.

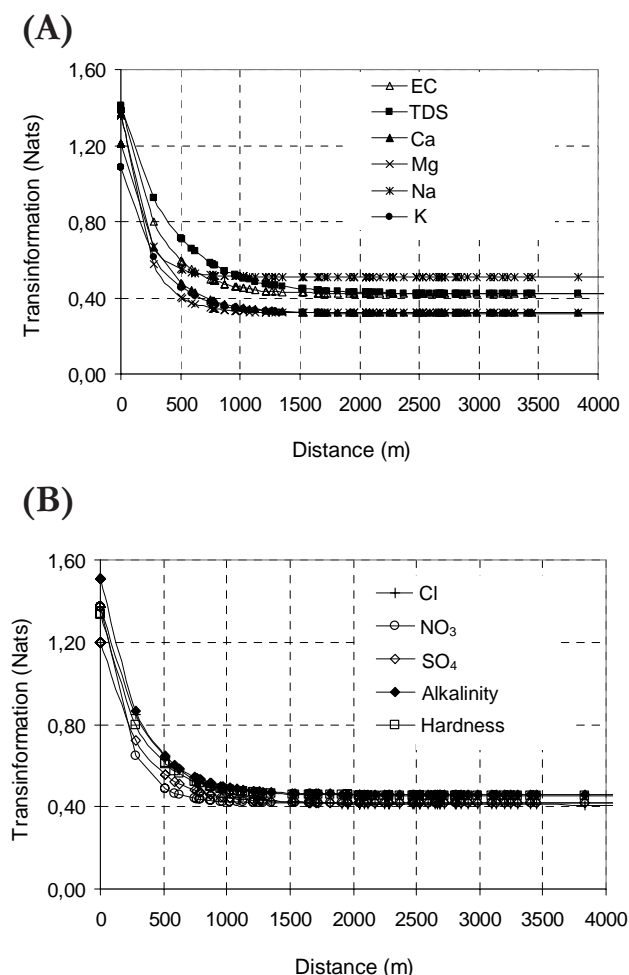


Fig. 9. Transinformation Model for groundwater quality variables: (A) EC, TDS, Ca, Mg, Na, and K; and (B) Cl, NO₃, SO₄, alkalinity, and hardness. The symbols (●, ♦, Δ, ... etc.) are used to distinguish between the eleven T Models of the different groundwater quality variables.

DETERMINING THE RANGE IN THE T MODEL (L)

The L value helps to estimate the degree of dependency between wells for one variable (spatial dependency). Moreover, it can also be used to reflect the dependency between the variables. The L value was estimated for all variables using Eqn. (10). In this equation, the values of T_0 , T_{\min} and K are taken from the best fitted exponential decay curve as represented in Fig. 9 and Table 3. Moreover, in Eqn. (10), the value of ϵ is considered as 0,0001 Nats which is particularly small. Under these conditions, the range L in the T Model can be estimated. Table 4 shows the values of the T Model parameters: T_0 , T_{\min} , K and the range L for all variables.

Table 4. TM_{ED} parameters for: EC, TDS, Ca, Mg, Na, K, Cl, NO₃, SO₄, alkalinity and hardness. ϵ is considered as 0,0001 Nats for computing L values.

Variables	T_0 (Nats)	T_{\min} (Nats)	K (1/m)	L (m)
EC	1,4100	0,4200	0,0034	2706
TDS	1,3570	0,4200	0,0024	3811
Ca	1,2100	0,3200	0,0034	2675
Mg	1,3700	0,3200	0,0050	1852
Na	1,3600	0,5100	0,0060	1508
K	1,0867	0,3167	0,0034	2632
Cl	1,3700	0,4100	0,0028	3275
NO ₃	1,3686	0,4186	0,0050	1796
SO ₄	1,1986	0,4186	0,0034	2636
Alkalinity	1,5100	0,4600	0,0034	2723
Hardness	1,3350	0,4550	0,0034	2671

SENSITIVITY ANALYSES

From Eqn. (10), it is inferred that the estimation of L is influenced by T_0 , T_{\min} , K and ϵ . Figure 11 shows the difference between T_0 and T_{\min} for all groundwater quality variables. It also represents the amount of redundant information, which decays as the distance increases. Generally, a higher amount of $(T_0 - T_{\min})$ requires a larger L value to be depleted. Estimation of the T_0 and T_{\min} values can be influenced by the number of class intervals. This is an essential element in the computation of the Transinformation values using the contingency tables (discrete approach).

In the T Model, the amount of redundant information ($T_0 - T_{\min}$) is diminished to reach a minimum value (very close to zero) with a decay rate K. Figure 12 shows that K is inversely proportional to the range L. This relation can

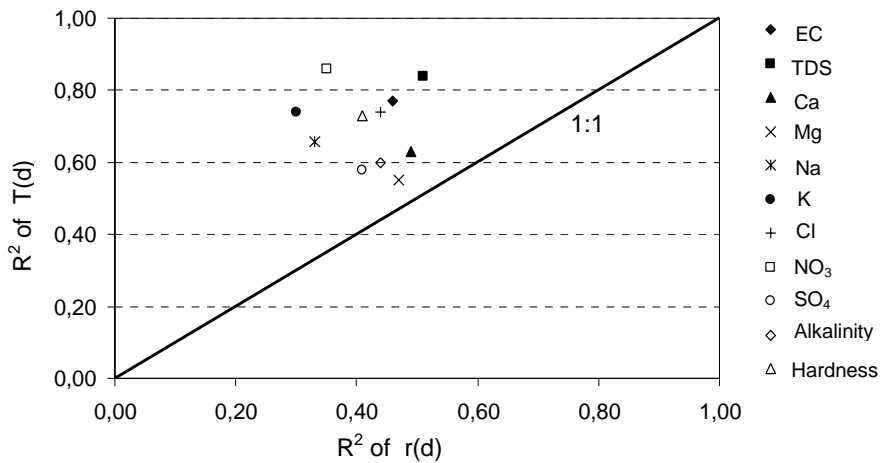


Fig. 10. Comparison of R^2 of the Transinformation model and the R^2 of the Correlation model, for all groundwater quality variables: EC, TDS, Ca, Mg, Na, K, Cl, NO_3 , SO_4 , alkalinity and hardness.

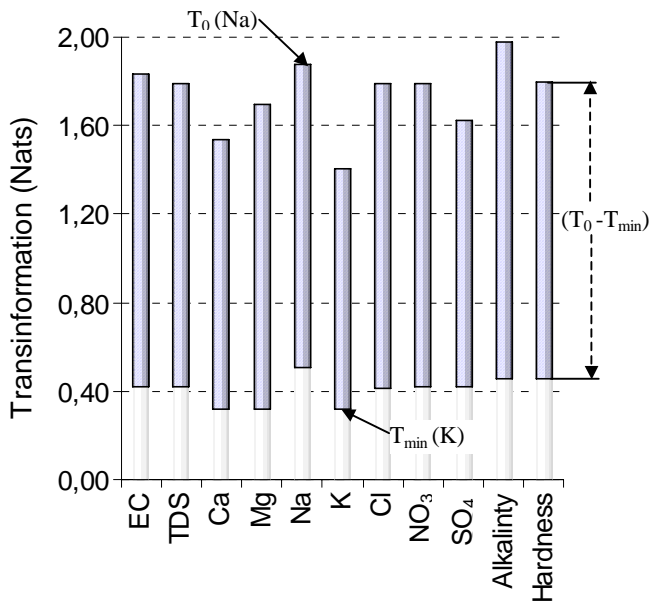


Fig. 11. The initial Transinformation (T_0) and the minimum Transinformation (T_{\min}) for all groundwater quality variables: EC, TDS, Ca, Mg, Na, K, Cl, NO_3 , SO_4 , alkalinity and hardness.

be expressed as: the amount of redundant information (Transinformation) which completely decays in a shorter distance when the decay rate of Transinformation (K) is high. Figure 13 shows also the effect of ϵ value on the estimated L value. The analyses of all groundwater variables revealed that increasing the ϵ value decreases the estimated L . Increasing the ϵ value means neglecting some redundant information between wells. Therefore, the value of ϵ should be small enough to ensure minimum redundant information.

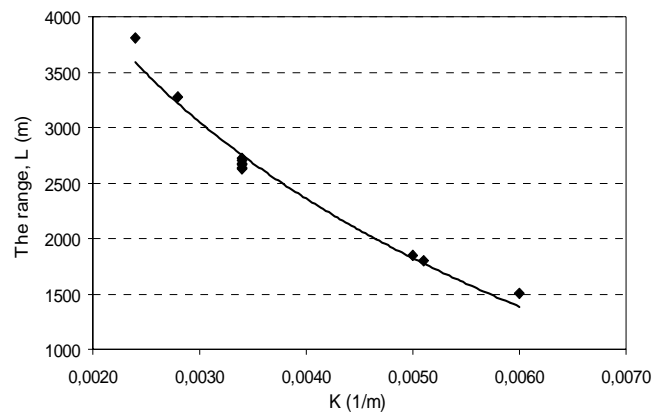


Fig. 12. The variation of the Transinformation decay rate (K) with the range (L) for all groundwater quality variables: EC, TDS, Ca, Mg, Na, K, Cl, NO_3 , SO_4 , alkalinity and hardness.

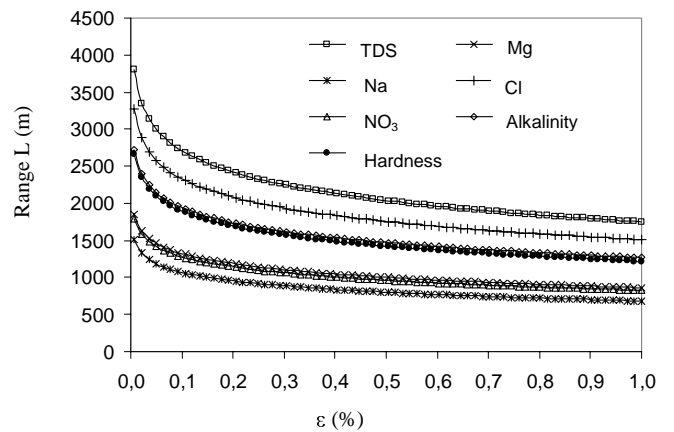


Fig. 13. The variation of the ϵ (%) with the range (L) for the groundwater quality variables: TDS, Mg, Na, Cl, NO_3 , alkalinity and hardness.

Discussion

From Fig. 9 and Table 4, parameters of the T Model, infer the spatial dependency between wells for one variable. In the analyses of the T Model of the groundwater quality variables, it was found that L and K dominate the description of spatial dependency between wells. For instance, the greatest dependency between wells can be found for TDS, since L is the highest among the variables (3811 m) and K is the lowest (0.0024). The lowest dependency can be found for Na, since L equals 1508 m and K is 0.0060. It also can be observed that EC, Ca, K, SO_4 , alkalinity and hardness have the same K value (0.0034), and L for these variables ranges between 2632 to 2723 m. In the T-Model of Cl, the L value equals 3275 m and K is 0.0028. Mg and NO_3 have L values of 1852 m and 1796 m, respectively, and the K value for both variables is 0.0050.

These evaluations may contribute to redesigning (reducing or extending) the groundwater quality monitoring network in the study area (Northern Governorate). For example, the number of wells to monitor TDS will be less than the number of wells to monitor Na. In the TDS case, the existing monitoring network can be reduced, since distances between adjacent wells can be less than 3811 m. However, in the Na case, the network can be extended by adding new wells, since there are distances between the adjacent wells greater than 1508 m. It can be understood that a complete monitoring network design methodology should be arranged by specifying: the objective of monitoring, the variable to be monitored, the strategy of monitoring (Transinformation, cost, hydrogeological behaviour, etc). That will certainly affect the number, location of monitoring wells (spatial) and monitoring temporal frequency.

The results of the T Model of the variables can also express the dependency between the variables. As indicated in Table 4, EC, Ca, K, SO_4 , alkalinity and hardness variables have similar L and K values. Thus, these variables can be classified in one group. Based on the same criteria Na, NO_3 and Mg can be classified in a second group, while TDS and Cl are in a third group. These classifications (clustering) can be used to reduce, eventually, the variables to be monitored. This will reduce the cost of monitoring, which is considered as an essential criterion for an optimal groundwater monitoring network design.

Conclusions

In this study, the groundwater quality in the Northern part of the Gaza Strip was analysed using the Transinformation Model (T Model) and Correlation Model (C Model). The analysis showed that for all the studied variables (EC, TDS,

Ca, Mg, Na, K, Cl, NO_3 , SO_4 , alkalinity and hardness), the T Model performed better than the C Model in characterising the spatial variability.

The T Model of the groundwater quality variables yielded different values for L, K and T_{\min} to be used for design purposes. A greater L reflects a higher dependency between wells for one variable and a smaller L indicates less dependency between wells. These results can be used for selecting the number and the locations of wells to be used in monitoring these variables. The results also revealed that some variables had similar T Model parameters (T_{\min} and L). This similarity can infer the degree of dependency between the variables and can also lead to a reduction of the variables to be monitored. Basically, the parameters of the T Model (T_{\min} and L) can be comparable to the semivariogram parameters (sill and range), which are commonly used to describe the spatial structure of groundwater variables.

For further steps toward the use of the entropy theory in groundwater monitoring network design, the methodology of this article will be utilised. All the groundwater quality monitoring wells in the Gaza Strip will be evaluated and redesigned according to the criteria mentioned in this paper. Optimisation techniques may be used for the design problem based on the criteria mentioned in this paper.

Acknowledgements

The authors gratefully thank the Foundation for Science and Technology of the Portuguese Ministry of Science and Technology for sponsoring the fellowship (Reference: SFRH/BD/6089/2001). The fellowship was provided for the first author's PhD programme entitled "Quantification of Information for Groundwater Quality Networks". The programme is being undertaken at the Department of Civil Engineering, Faculty of Science and Technology, University of Coimbra, Coimbra, Portugal.

References

- Ahmed, S., 2002. Groundwater monitoring network design: applications of Geostatistics with a few case studies from a granitic aquifer in a semi-arid region. In: *Groundwater Hydrology*, M.M. Sherif, V.P. Singh and M. Al-Rashed (Eds.), Balkema, Tokyo, Japan. **2**, 37–57.
- Bueso, M.C., Angulo, J.M. and Alonso, F.J., 1998. A state-space model approach to optimum spatial sampling design based on entropy. *Environ. Ecol. Stat.*, **5**, 29–44.
- Cressie, N., 1990. *Statistics for Spatial Data. Revised Edition*. Wiley, New York, USA. 900pp.
- Geo, H., Wang, J. and Zhao, P., 1996. The update Kriging variance and optimal sampling design. *Math. Geol.*, **28**, 295–313.
- Gokhale, D. and Kullback, S., 1978. *The information in Contingency Tables*. Marcel Dekker Inc., New York, USA. 365pp.

- Golden Software, 1999. *Surfer Version 7, Surface Mapping System*. Golden Software Inc., Colorado, USA. www.goldsoftware.com.
- Harmancioglu, N.B., Fistikoglu, O., Ozkul, S.D., Singh, V.P. and Alpaslan, N., 1999. *Water Quality Monitoring Network Design*. Kluwer, Boston, USA. 299pp.
- Hudak, F. and Loaiciga, H., 1993. An optimisation method for monitoring network design in multilayered groundwater flow system. *Water Resour. Res.*, **29**, 2835–2845.
- Husain, T., 1989. Hydrologic uncertainty measure and network design. *Water Resour. Bull.*, **25**, 527–534.
- Jager, H.I., Sale, M.J. and Schmayer, R.L., 1990. Cokriging to assess regional stream quality in the Southern Blue Ridge Province. *Water Resour. Res.*, **26**, 1401–1412.
- Journel, A. and Huijbregts, C.H., 1978. *Mining Geostatistics*. Academic Press Inc., New York, USA. 600pp.
- Kapur, J.N. and Kesavan, H.K., 1992. *Entropy Optimisation Principles with Applications*. Academic Press Inc., San Diego, USA. 408pp.
- Krstanovic, P.F. and Singh, V.P., 1992. Evaluation of rainfall networks using entropy: II. applications. *Water Resour. Manage.*, **6**, 295–314.
- Loaiciga, H., 1989. An optimisation approach for groundwater quality monitoring network design. *Water Resour. Res.*, **25**, 1771–1782.
- Lubbe, C.A., 1996. *Information Theory*. Cambridge University Press, Cambridge, UK. 350pp.
- Marsily, G., 1986. *Quantitative Hydrogeology*. Academic Press Inc., New York, USA. 435pp.
- Metcalf and Eddy, 2000. *Costal Aquifer Management Program, Final Report: Modelling of Gaza Strip Aquifer*. The program is funded by US Agency for International Development (USAID) and owned by the Palestinian Water Authority (PWA). Gaza, Palestine.
- Mogheir, Y. and Singh, V.P., (2002a). Application of information theory to groundwater quality monitoring networks. *Water Resour. Manage.* **16**, 37–49.
- Mogheir, Y. and Singh, V.P., (2002b). Specification of information needs for groundwater resources management and planning in developing country: Gaza Strip case study. In: *Groundwater Hydrology*, M.M. Sherif, V.P. Singh and M. Al-Rashed (Eds.), A.A. Balkema, Publishers, Tokyo, Japan. **2**, 3–20.
- Mogheir, Y., de Lima, J.L.M.P. and Singh, V.P., (2004a). Characterizing the spatial variability of groundwater quality using the entropy theory: I. Synthetic data. *Hydrol. Process.*, (in press).
- Mogheir, Y., de Lima, J.L.M.P. and Singh, V.P., (2004b). Characterizing the spatial variability of groundwater quality using the entropy theory: II. Case study from Gaza Strip. *Hydrol. Process.*, (in press).
- Motulsky, H.J., 1999a. *Analysing Data with GraphPad Prism3, A companion to GraphPad Prism version 3*. San Diego, USA. www.graphpad.com.
- Motulsky, H.J., 1999b. *Analysing Data with GraphPad Prism3, GraphPad Software*. San Diego, USA, www.graphpad.com.
- Olea, R.A., 1984. Sampling design optimisation for spatial functions. *Math. Geo.*, **16**, 369–393.
- Olea, R.A., 1999. *Geostatistics for Engineers and Earth Scientists*. Kluwer Academic Publishers, Boston.
- Rouhani, S. and Hall, T.J., 1988. Geostatistical schemes for groundwater sampling. *J. Hydrol.*, **103**, 85–102.
- Sounga, J.O., 1976. Entropy principle applied to the rainfall-runoff process. *J. Hydrol.*, **30**, 81–94.
- William, J.D., 1997. *Groundwater Geochemistry: Fundamentals and Application to Contamination*. Boca Raton Lewis, New York, USA. 221pp.
- Wu, S. and Zidek, V.J., 1992. An entropy-based analysis of data from selected NADP/NTN network sites for 1983–1986. *Atmos. Environ.*, **26A**, 2089–2103.
- Yfantis, E.A., Flatman, G.T. and Behar, J.V., 1987. Efficiency of Kriging estimation for square, triangular and hexagonal grids. *Math. Geol.*, **19**, 183–205.
- Zhou Y., 1996. *Spatial Data Generation Program (COVRAN)*. Delft, The Netherlands.